Pulmonary Sound Analysis with Deep Learning for Efficient Respiratory Disease Categorization

V. Ricky Deeven(✉), V. Naveen Kumar, Y. Padma Sai, N. Akshitha, and M. Kaivalya

Department of Electronics and Communication Engineering, VNRVJIET, Hyderabad, Telangana, India
rickydeeven2411@gmail.com, padmasai_y@vnrvjiet.in

Abstract. Lung Disease treatment is crucial in the medical industry since it is the third most prevalent cause of mortality worldwide. Recently, assistive solutions have greatly benefited from the use of technologies like deep learning and machine learning. This involves the use of a variety of technologies, including Magnetic resonance imaging (MRI), isotopes, X-rays, and CT scans. Unfortunately, using auscultation to identify these disorders requires qualified doctors, and this approach is insufficiently objective. So, it is essential to have a mechanism for accurately recognizing. The proposed method successfully converts recorded audio signals into spectrogram images utilizing the time-frequency approach and deep convolutional neural network (CNN) with prior training which achieved a 94% validation accuracy. The primary objective of this project is to identify healthy lungs and respiratory conditions using ICBHI 2017 data set. As this project considered all seven classes of respiratory sounds at once, the findings based on this framework are superior to those obtained using earlier methods.

Keywords: Respiratory sounds · Pulmonary sounds · Convolutional neural network (CNN) · Mel spectrogram

1 Introduction

Lung sounds can indicate many things about your respiratory health and disease. Making sure that you hear your lung sounds can help you determine if there are abnormalities in your breathing. The air movements and changes in the lung tissue that occur during breathing are what cause the noises that a person makes when they breathe. The two basic kinds of lung sounds are normal lung sounds and aberrant lung sounds. While crackles and wheezes are considered problematic lung sounds [1], bronchial and vesicular noises are considered typical lung sounds. Respiratory sound disorders can be identified based on lung sounds.

Respiratory disorders are a leading cause of sickness worldwide, causing over 4 million premature deaths annually, second only to cardiovascular disorders. Currently, CT scans and chest X-rays are commonly used to diagnose these disorders, but they are expensive and emit radiation. Listening to the lungs with a stethoscope is helpful, but not conclusive. Machine learning (ML) could provide a more accessible and accurate

B. Raj et al. (Eds.): ICETE 2023, AER 223, pp. 68–78, 2023.
https://doi.org/10.2991/978-94-6463-252-1_9

diagnostic tool, especially in rural or undeveloped areas lacking resources. Adventitious sounds, such as crackles and wheezes, detected during lung auscultation, are crucial in the diagnosis of respiratory disorders. Automating the identification of these sounds through ML can facilitate the early detection of these disorders.

Researchers in [1] and [3] employed convolutional neural networks (CNNs) to classify respiratory sounds, and the CNN model surpassed standard machine learning approaches in terms of accuracy. Abnormal breath sounds include low-pitched rhonchi caused by blockages in big airways and high-pitched crackles associated with diseases such as pneumonia, bronchitis, and asthma. Stridors and wheezes are also sounded that indicate respiratory conditions and cough patterns can be used to analyze lung sounds for conditions such as widened or narrowed airways, fluid-filled air sacs, and stiff lungs. Studies have detected crackles and wheezes in lung sounds using the short-time Fourier transform spectrogram. Here CNN [14] and MFCC [2] were used in another with reduced accuracy. However, no research has classified respiratory illnesses based on lung sounds, such as bronchiectasis, pneumonia, bronchiolitis, COPD, URTI, and healthy.

Many recent studies have explored accurate and reliable multi-class classification of respiratory diseases based on lung sounds. The relationship between cough patterns and respiratory conditions, such as widened airways, a narrowed airway, fluid-filled air sacs, and stiff lungs, can be established in [4], same can be applied to analyzing the lung sounds.

The MLNN approach [5] used single and double hidden layers with the Levenberg–Marquardt (LM) algorithm, but no deeper layers were explored for potentially better results. Mel spectrograms were used for convolutional networks to classify respiratory audios, with a Variational Convolutional Autoencoder (VAE) [6] proposed to improve unusual classes. However, more sample and training data are needed for more accurate results and to determine the model's true accuracy. The studies [7] demonstrate the potential of machine learning and deep learning approaches for accurate and reliable multi-class classification of lung sounds.

According to this study [8], K closest neighbor (KNN), decision trees, artificial neural networks (ANN), and SVM were determined to be the most suitable among several classification algorithms when it came to feature selection. Nevertheless, it does not differentiate between mild, moderate, and severe illness severity levels. One of the key challenges in this field is to extract informative features from lung sounds that can be used to classify different disease categories. In the study by Rocha [9], the authors used the ICBHI 2017 challenge database for multi-class classification of normal, wheezes, crackles, and combination wheezes.

Due to the inability to hear lower frequency waves, physician expertise, and the surrounding environment, the manual stethoscope may result in misdiagnosis or undiagnosed illnesses. So, using electronic stethoscopes [10], as a substitute, can save lung sounds as signals within a computer, allowing time-frequency analysis for better and more accurate interpretation by medical personnel.

Table 1. Respiratory Cycle dataset

Disease	No. of patients
Bronchiectasis	16
Pneumonia	37
COPD	793
Bronchiolitis	13
URTI	23
Healthy	25

2 Materials and Methods

This framework involves acquiring respiratory audio sounds, standardizing their length, converting them to Mel spectrogram images, and classifying them into different categories using a convolutional neural network. The model is built using Keras and Tensorflow, and its accuracy is assessed through a signal test, which shows better results than previous methods.

2.1 Data Acquisition

Data for the research was collected from the ICBHI challenge and includes 920 annotated audio samples from 126 participants with both healthy and respiratory illness conditions. Table 1 presents the number of patients for each class of disease. Several stethoscopes, including the AKGC417L, Meditron, Litt3200, and LittC2SE, were used to record the audio. The rate of sampling extends from 4kHz to 44kHz, and the keeping time spans 10s to 90s.

Respiratory sounds of both healthy people and those with respiratory illnesses are included in the data samples. All age categories, including infants, teens, adults, and the elderly, are represented among the patients. The collection consists of 920 annotated audio samples from 126 participants that span 5.5 h and 1,864 out of the 6898 respiratory cycles have crackles, 886 have wheezes, and 506 have both features.

2.2 Data Preprocessing and Feature Extraction

The dataset used in the study contained unstructured data and anomalies. To standardize the data. We used the Python module Librosa to trim and pad to a length of 20 s of the audio files to standardize the data. The following describes the deep learning pipeline's preprocessing stage. The acoustic files from the LRTI and asthma patients were first deleted. This was due to the lack of patients and audio data in those classes. Each sound file was loaded into memory using a sampling rate of 11,025 and divided into 512-sample segments, which were converted into 40 MFCCs with frequencies ranging from 50 to 2k. As a result, the 917 sound files produced an MFCC of size 431x40.

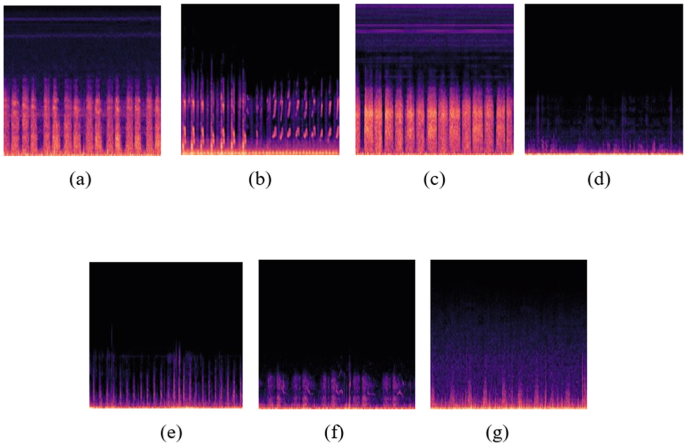


Fig. 1. Spectrogram images of respective diseases

Figure 1 represents the spectrogram images of the diseases and healthy lungs where (a) is Bronchiectasis, (b) is COPD, (c) is URTI, (d) is asthma, (e) is LRTI, (f) is pneumonia and (g) is Healthy. The audio recordings were thereafter subjected to speed augmentation, and during data augmentation, iterations over each sound file were conducted to extract features using MFCC. The Neural network was given these spectrogram characteristics for further categorization. By taking time-varying air pressure samples and performing a quick Fourier Transform to convert them into the frequency domain and subsequently produced the Mel-Spectrogram by converting the frequency into a Mel scale and the color dimension into the amplitude. The sound's transient power spectrum was used. The audio recordings underwent speed augmentation, and features were extracted using MFCC through iterations over each sound file during data augmentation. The Neural network categorized the recordings based on these spectrogram features.

2.3 Model Explanation

The CNN classifier uses convolution to extract features, followed by MaxPooling2D to optimize specifications and reduce overfitting risks. GlobalAveragePooling2D is used in the final layer, and SoftMax is used to activate the output layer and interpret the output number as a probability (Fig. 2).

The model architecture consists of multiple layers with varying numbers of filters and kernel sizes. The first layer has 64 kernels of size 3 and uses ReLU activation. The input layer has dimensions of 63 x 64 x3 and a max-pooling stride of 2. The second layer has 128 filters of size 3 x 3 and uses ReLU activation. The third layer has 256 filters with a 3x3 kernel and no shift invariance mechanism. A Dense-1 layer with 512 hidden units is added, followed by a 30% dropout to combat variation. The last layer (Dense-2) uses 'soft argmax' as a transfer function for multi-class classification. Padding is used in the second layer for Soft and Log Mel spectrum, and "valid" padding is used for MFCC feature extraction.

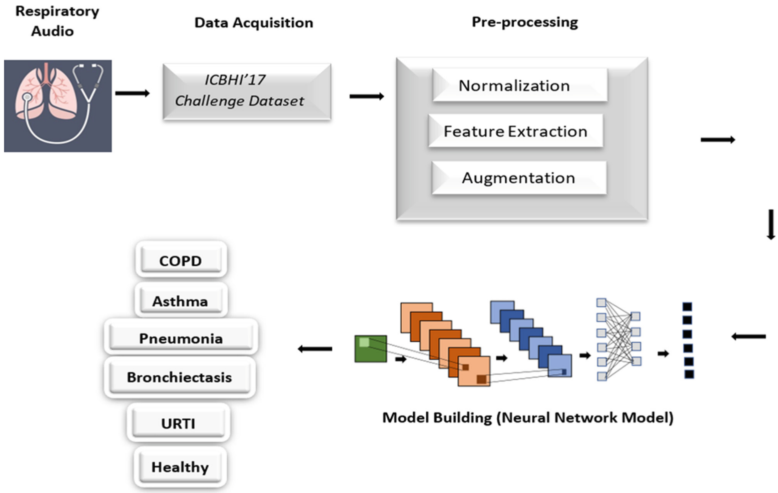


Fig. 2. Schematic diagram for model explanation

The 'flatten' layer is used to convert the 2-Dimensional arrays into a single linear vector. The output layer has six neurons representing bronchiectasis, pneumonia, bronchiolitis, COPD, URTI, and healthy. The suggested model's CNN architecture is presented in Table 2. The noise level in audio inputs is often higher than in image inputs. The Adam algorithm, which combines stochastic gradient descent and the RMSprop method, provides better weight optimization logic and hyperparameter tuning. It offers faster convergence and optimized performance compared to other optimizers like SGD and ADAGrad.

3 Results and Discussion

3.1 Performance Analysis

The confusion matrix was employed (see Fig. 3) for purposes of assessment and analysis. The accuracy can be computed by the number of accurate forecasts divided by the total number of predictions given by-

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ positive + True\ Negative + False\ positive + False\ Negative} \tag{1}$$

Precision and recall are provided as metrics to assess the model's effectiveness. To account for both FPs and FNs, the F1-Score often combines accuracy and recall into a single metric. It comes from –

$$Precision = \frac{True\ Positive}{True\ Positive + False\ positive} \tag{2}$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{3}$$

Table 2. The architecture of the proposed model

Layers	Shape (Output)	Parameters
Convolution_2D (1)	(None, 62, 62, 32)	896
MaxPooling_2D (1)	(None, 31, 31, 32)	0
Convolution_2D (2)	(None, 29, 29, 64)	18496
MaxPooling_2D (2)	(None, 14, 14, 64)	0
Convolution_2D (3)	(None, 12, 12, 64)	36928
MaxPooling_2D (3)	(None, 6, 6, 64)	0
Convolution_2D (4)	(None, 4, 4, 32)	18464
MaxPooling_2D (4)	(None, 2, 2, 32)	0
Dropout	(None, 2, 2, 32)	0
Flatten	(None, 128)	0
Dense (1)	(None, 128)	16512
Dense (2)	(None, 7)	903
		92,199

$$F1 - score = \frac{2True\ Positive}{2True\ Positive + False\ Negative + False\ Positive} \tag{4}$$

The dataset of 126 individuals was split into training and validation sets, with 80% used for training and 20% for validation. The model's performance was evaluated using

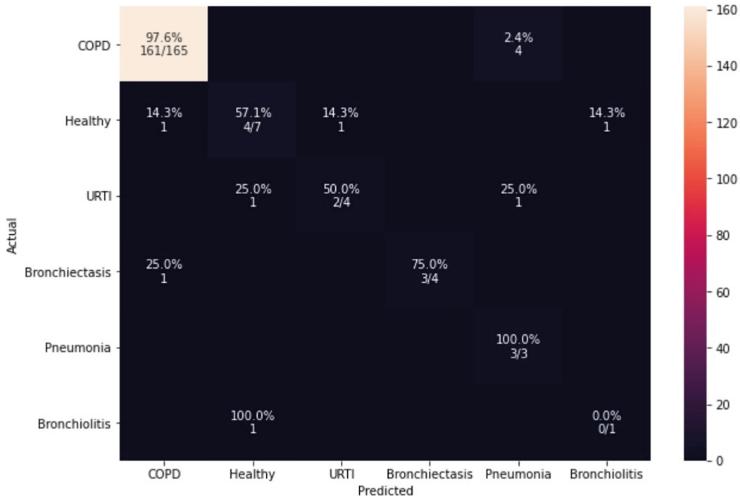


Fig. 3. Confusion Matrix

a confusion matrix, with precision and recall calculated for each type of lung sound. The accuracy, recall, and f1-score are determined using the confusion matrix in Table 3.

Using spectrogram pictures and categorizing them into three groups, they were able to achieve an accuracy of 87% using VGG - BDGRU, as demonstrated in [10]. Even though we just employed 2D CNN, we outperformed them with an accuracy of 94.26% and categorized them into 6 classes. The first study used an MLNN [11] with a damped least squared method with one and another with two hidden layers, achieving accuracies ranging from 90% to 95.43%. The double hidden layer neural network was found to be preferred, but further layers were not explored. The second study used a hybrid neural model [12] with CNN and LSTM, achieving accuracies of 73.69% and 76.39%. This proposed model using 200 batches and 70 epochs optimized with the Adam approach achieved higher accuracy and required less time for training.

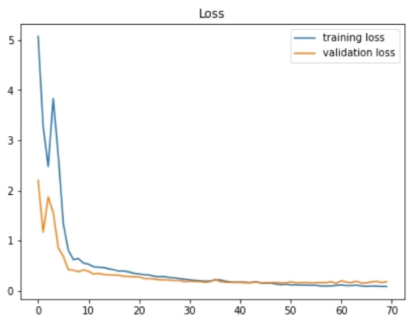
Figure 4 (a) displays the training and validation accuracy curves, with training accuracy at 0.9741 and validation accuracy at 0.94. Figure 4 (b) shows the corresponding loss curves, with training loss at 0.0870 and validation loss at 0.1827.

Table 4 lists the CNN model's training parameters and accuracy. In this project, 200 batches were allocated, and 70 epochs were set as the default to avoid any over- or

Table 3. Classification Report

	Precision	Recall	F1-score	Support
COPD	0.99	0.98	0.98	165
Healthy	0.67	0.57	0.62	7
URTI	0.67	0.50	0.57	4
Bronchiectasis	1.00	0.75	0.86	4
Pneumonia	0.38	1.00	0.55	3
Bronchiolitis	0.00	0.00	0.00	1
Accuracy			0.94	184

(a)



(b)

Fig. 4. Accuracy curve and Loss curve

Table 4. Accuracy and other elements

Optimizer	Epochs	Sample Size	Training Time (minutes)	Accuracy
Adam	70	200	6 min	94%

under-fitting and proposed model was optimized using the Adam approach since it is computationally effective, uses less memory, and performs better with noisy input data. Based on the epochs and Sample size, the model's overall accuracy was determined.

3.2 Comparison of Existing Work and Current Study

Table 5 compares our proposed model with related works and demonstrates that our model achieves the highest accuracy of 94%. While the model presented in [1] achieved accuracy equal to the proposed model, it was limited by the availability of data for certain illnesses such as bronchiectasis and pneumonia. We used innovative techniques such as data augmentation and normalization to preprocess the data and remove unwanted noises from other body parts. The present study outperformed similar research in the categorization of respiratory disorders.

Table 5. Comparison of related works

Study	Model	No. patients	No. recordings	No. classes	Accuracy
-	Proposed Model	126	6896	Normal, Asthma, Bronchiectasis, Pneumonia, COPD, URTI (6)	94%
[1]	CNN	15	2140	Normal, monophonic wheeze, polyphonic wheeze, stridor, fine crackle, squawk, coarse crackle (7)	94%
[2]	CNN + LSTM	126	6896	Normal, wheeze, crackles, and both (4)	74.3%
[3]	SVM/CNN	1630	15,328	Normal, rale rhonchus (3)	80%
[5]	SVM	155	-	COPD, Healthy (2)	65%
[8]	KNN	50	50	COPD, Normal (2)	93%
[10]	VGG - BDGRU	384	1152	Normal, asthma, pneumonia (3)	87%
[13]	CNN	126	6896	Normal, crackles, wheezes, crackles + wheezes (4)	71.5%

4 Conclusion

In this paper, straightforward and less demanding resources on CNN-based deep learning assistant methods help professionals in medicine to identify lung disorders using respiratory sounds. In comparison to many other models already in use, the proposed model's average accuracy in classifying lung sounds into different respiratory states was 94% for the tabletop position. The implementation of models taught by deep learning in clinical contexts to support physician decision-making is made possible by this work.

The capabilities of a machine learning model can be enhanced to aid doctors in detecting various medical conditions, including the probability of a heart attack based on the sound of heartbeats. Strategies such as data augmentation, deep learning, and feature normalization may be utilized to enhance performance. Additionally, future initiatives may focus on pinpointing the exact region of lung disease diagnosis in audio sessions. While initial goals have been met, further work is required to improve the model's precision before it can be used in hospitals.

References

1. Bardou D, Zhang K, Ahmad SM. Lung sounds classification using convolutional neural networks. Artif Intell Med., May 2018.
2. A.A.Saraiva, D. B. S. Santos, A. A. Francisco, Jose Vigno Moura Sousa, Antonio Valente, “Classification of respiratory sounds using convolutional neural network,” 11th International Conference on Bioinformatics Models, Methods and Algorithms, pp. 138–144, January 2020.
3. Aykanat, M., Kilic, Ö., Kurt, B. et al. Classification of lung sounds using convolutional neural networks. J Image Video Proc. 2017, 65, Aug 2017.
4. Rudraraju, Gowrisree & Palreddy, ShubhaDeepti & Mamidgi, Baswaraj & Sripada, Narayana Rao & Padma Sai, Y. & Kumar, V.Naveen & Haranath, Sai, Cough sound analysis and objective correlation with spirometry and clinical diagnosis. Informatics in Medicine Unlocked, March 2020
5. Er, Orhan & Temurtas, Ullah, “A Study on Chronic Obstructive Pulmonary Disease Diagnosis Using Multilayer Neural Networks.” Journal of medical systems, 32. 429–32, 2018.
6. García-Ordás, María Teresa, Carmen Benavides, “Detecting Respiratory Pathologies Using Convolutional Neural Networks and Variational Autoencoders for Unbalancing Data” Sensors 20, no. 4: 1214, 2020.
7. Aykanat, Murat & Kilic, Ozkan & Kurt, Bahar & Saryal, Sevgi. (2017). “Classification of lung sounds using convolutional neural networks.” EURASIP Journal on Image and Video Processing, 2017.
8. Amaral, Jorge L M et al. “Machine learning algorithms and forced oscillation measurements applied to the automatic identification of chronic obstructive pulmonary disease.” Computer methods and programs in biomedicine vol. 105,3, 2012.
9. Rocha BM, Filos D, Mendes L, Serbes G, Ulukaya S, Kahya YP, “An open access database for the evaluation of respiratory sound classification algorithms.” Physiological measurement vol. 40,3 035001. 22 March 2019.
10. L. Shi, K. Du, C. Zhang, H. Ma and W. Yan, “Lung Sound Recognition Algorithm Based on VGGish-BiGRU ,” in IEEE Access, vol. 7, pp. 139438–139449, Sept 2019.
11. Habukawa C, Ohgami N, Matsumoto N, “A wheeze recognition algorithm for practical implementation for practical implementation in children,” PloS one vol. 15,10, October 2020.
12. Q. Chen, W. Zhang, X. Tian, X. Zhang, S. Chen, and W. Lei, “Automatic heart and lung sounds classification using convolutional neural networks”, Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), pp. 1–4, 2016.
13. Demir, F., Sengur, A. & Bajaj, V, Convolutional neural networks based efficient approach for classification of lung diseases, Health Inf Sci Syst 8, 4 , Dec 2020.
14. Rahib H. Abiyev, Mohammad Khaleel Sallam Ma’aitah, “Deep Convolutional Neural Networks for Chest Diseases Detection”, Journal of Healthcare Engineering, vol. 2018, 4168538, 11 pages, 2018.